NASA's Information Power Grid

William E. Johnston, Project Manager
Arsi Vaziri, Deputy Project Manager
Tom Hinke, Deployment Project
Tony Lissota, Implementation Manager
Piyush Mehrotra, Application Frameworks Technologies

NASA Advanced Supercomputing (NAS) Division
NASA Ames Research Center
John Ziebarth, Division Chief
William Thigpen, Engineering Branch Chief,

http://www.ipg.nasa.gov

What Are Computing and Data Grids?

- Grids are technology and an emerging architecture that involve several types of *middleware that mediate between science portals, applications, and the underlying resources* (compute, data, and instrument) in order to simplify the construction of large-scale problem solving systems
- *Grids are persistent environments* that facilitate integrating software applications with instruments, displays, computational, and information resources that are managed by diverse organizations in widespread locations

What Are Computing and Data Grids?

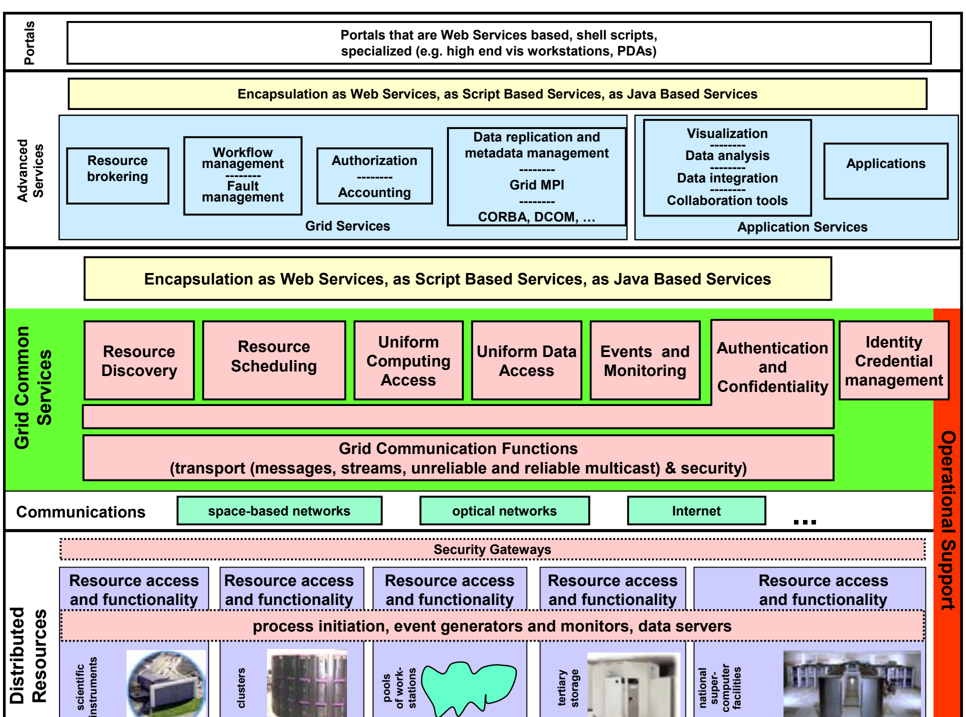
- *Grids are tools for data intensive science* that facilitate remote access to vary large amounts of data that is managed in remote storage resources and analyzed by remote compute resources, all of which are integrated into the scientist's software environment.
- Grids are persistent environments and tools to facilitate *large-scale collaboration* among global collaborators.
- Grids provide for securely sharing resources among institutional collaborators.

What Are Computing and Data Grids?

- Grids are also a *major international technology initiative* with 450 people from 35 countries in an IETF-like standards organization: The Global Grid Forum

Grids as the Future of Computing?

- *Grids and their integration with "Web Services"* may well be the general direction for business and other access to computing and information resources
- Web services are focused on managing services – instances of applications
- Grids are focused on managenign collections of resources
- There is a growing opinion in the commercial world that *Grids and Web Services* combined represent as powerful a new tool for accessing and managing distributed resources as the World Wide Web has proven to be for information distribution
- Note recent announcements from IBM and other mainstream computing and software suppliers and their strong participation in Global Grid Forum
  - IBM is well on its way to adopting a Web Grid Services architecture ("Websphere") as it's corporate architecture strategy because of Web Grid Services' potential for integrating legacy applications and making them all available through Web interfaces. Ditto Microsoft and .NET.



Why Are Grids Important?

- *Grids are being adopted and developed in several scientific disciplines that have to deal with large-scale collaboration, massive distributed data, and distributed computing problems*, e.g.
  - High Energy Physics (major US, Europe, and Asia Pacific Grids)
  - Observational astronomy and astrophysics (US, Europe, and Asia Pacific)
  - Earthquake engineering community (NEESGrid - US)
- *Grids are being deployed as infrastructure for science*
  - NASA's IPG
  - UK eScience Grid
  - EU Data Grid (Europe – actually many Grids)
  - DOE Science Grid (US – actually many Grids)
  - NSF TeraGrid (interconnected supercomputer Grid)
- Grids can facilitate NASA's large-scale science, engineering, and (maybe) operational systems

Points to Take Away

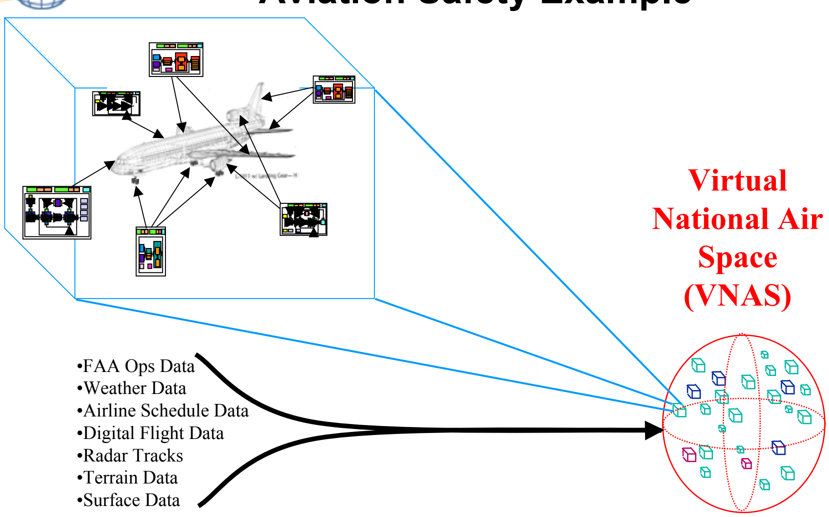
1. Grids are a major international effort to define a common computing, data management, and collaboration infrastructure for science
2. NASA has a major role in Grids
3. NASA is heavily leveraging a lot of other work in Grids

NASA's Information Power Grid

- NASA's large-scale science and engineering problems *require using many compute and data resources*, including supercomputers and large-scale data storage systems, *all of which must be integrated with applications and data* that are
  - developed by different teams of researchers
  - or that are obtained from different instruments

  and all of which are at different geographic locations.

Multi-disciplinary Simulations: Aviation Safety Example



The vision for VNAS is that whole system simulated aircraft are inserted into a realistic environment. This requires integrating many types of operations data as drivers for the simulations.

Information Power Grid Goal

- A persistent Computing and Data Grid that provides a *uniform environment for building and using large-scale, dynamically constructed problem solving environments* built from from distributed, heterogeneous resources.

State of Grids

- Grids are real, and they are useful now
- Basic Grid services are being deployed to support uniform and secure access to computing, data, and instrument systems that are distributed across organizations
  - resource discovery
  - uniform access to geographically and organizationally dispersed computing and data resources
  - job management
  - security, including single sign-on (users authenticate once for access to all authorized resources)
  - secure inter-process communication
  - Grid system management
  - Global events publish/subscribe (prototype)

State of Grids

- Higher level services
  - Grid execution management tools (e.g. Condor-G) are being deployed
  - Data services providing uniform access to tertiary storage systems and global metadata catalogues (e.g. GridFTP and SRB/MCAT) are being deployed
  - Web services supporting application frameworks and science portals are being prototyped

State of Grids

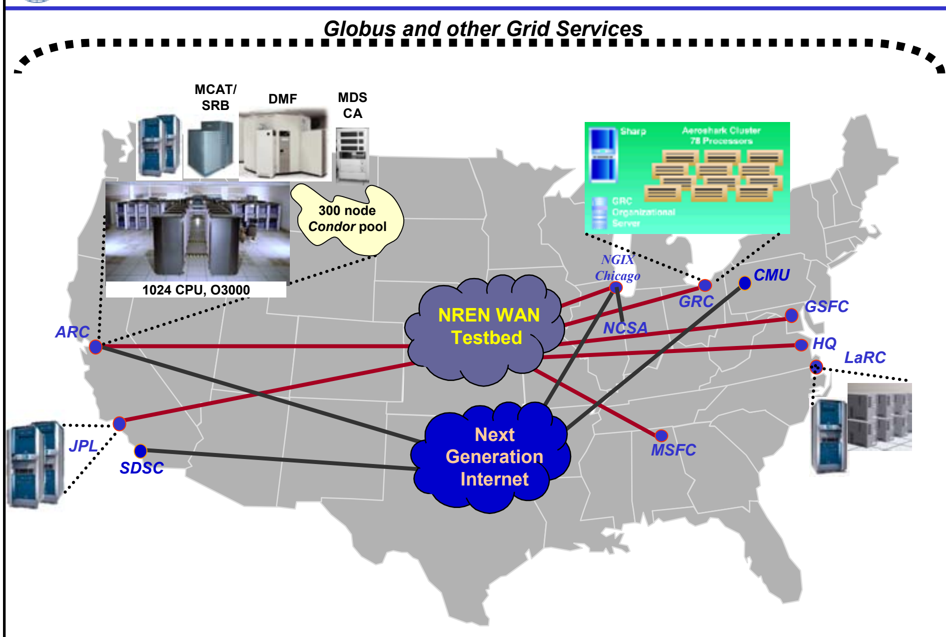
- Persistent infrastructure is being built
  - Grid services are being maintained on the compute and data systems in prototype production Grids
  - Cryptographic authentication supporting single sign-on is being provided through Public Key Infrastructure
  - Resource discovery services are being maintained (Grid Information Service – a distributed directory service)

State of the Information Power Grid

- Hardware resources for the baseline IPG prototype-production Grid system include
  - approximately 1500 CPU nodes in half a dozen SGI Origins distributed across several NASA centers
  - 10-50 Terabytes of uniformly and securely accessible mass storage
  - several workstation clusters involving about 100 CPUs
  - a Condor pool of about 300 workstations

  All of these are managed and accessed through the IPG Grid services
  - Globus (basic Grid functions)
  - SRB/MCAT (uniform access to distributed tertiary storage and global metadata catalogues)
  - Condor (pools of workstations as a Grid resource)

Information Power Grid, 2002



The State of IPG

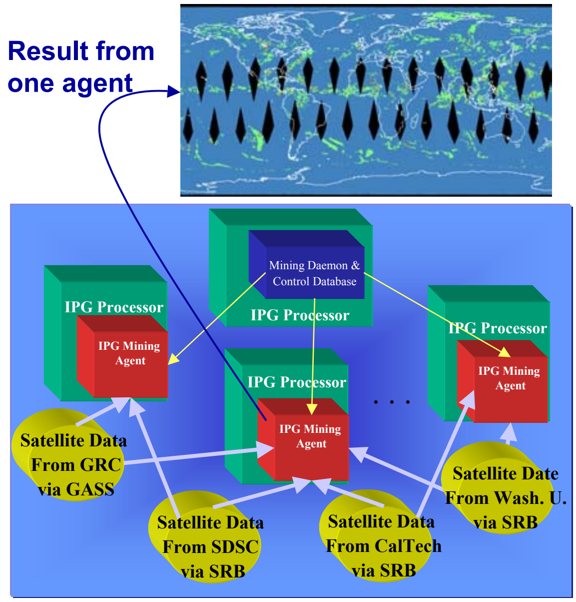
- IPG and NAS are *building and operating Grid infrastructure*, and *developing and deploying Grid services* for NASA's persistent Grid
- Several major milestones have demonstrated IPG as an operational, prototype-production Grid
- Several of these milestones have demonstrated that the *NASA Grid can interoperate with the other extant Grids* at Universities and other Federal Labs.
- This represents significant progress toward a "universal," common computing and data infrastructure for science

The Data Mining Using Grid Services

- Management and access to massive data sets is fundamental to large-scale science and engineering. The IPG Data Mining application demonstrated:
  - Persistent and uniform access to heterogeneous, multi-organizational archival storage systems
  - The SDSC Storage Resource Broker (SRB - an IPG Grid service) provides a standard data access interface for heterogeneous data archive systems
  - SRB's MCAT is a catalogue service that provides a standard way to define, manage, and search metadata for all files in a collection, where a collection may span many data archive systems – i.e. it provides for federating dataset in different systems

High Speed Distributed Data Access: IPG Milestone Completed 3/2000

- **Data access capabilities of IPG are demonstrated by *parallel data mining***
- **512 node SGI Origin at Ames uses IPG uniform interface data access tools (SRB) to simultaneously mine hydrology data from four sites**
  - **SDSC**
  - **CalTech**
  - **GRC**
  - **Washington U.**



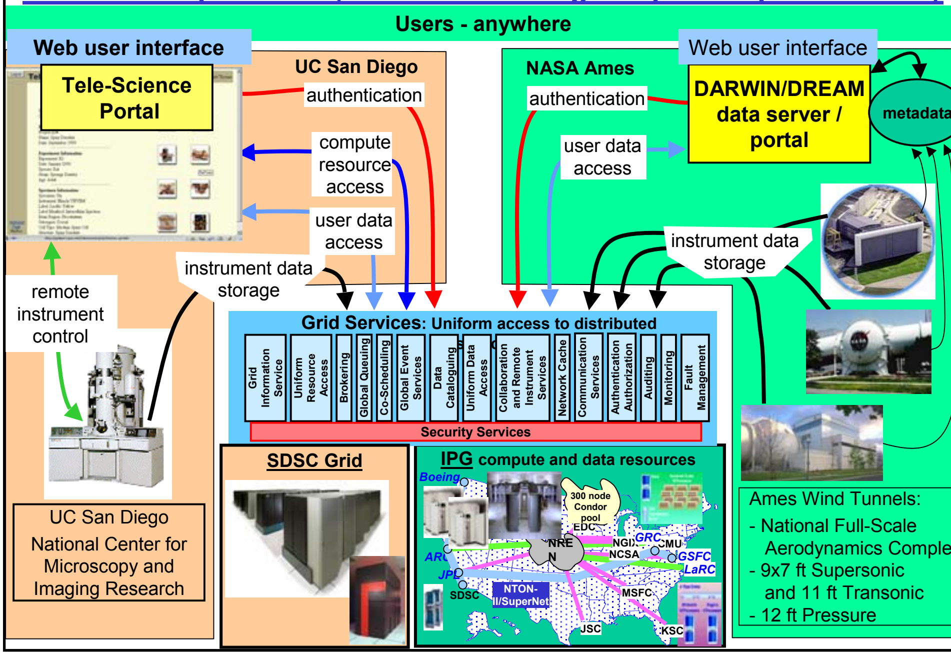
Tom Hinke, NASA Ames

Remote Access to High Data-Rate Instruments

- Two Grid based remote instrument systems have been demonstrated, one at Ames and one remote from Ames. Both use Grid services to provide secure data management and remote access to various aspects of the instruments. The DARWIN system at Ames has users scattered across the country, the UCSD TeleScience system has a NASA user at Wallops Island manipulating the instrument at UCSD. Both systems store data on IPG data resources at Ames. All of the critical data paths for both demonstrations transferred data at 50 Mb/s, or greater.
- This was an IPG 4QFY01, Level-1 Milestone

Remote Access to High Data-Rate Instruments

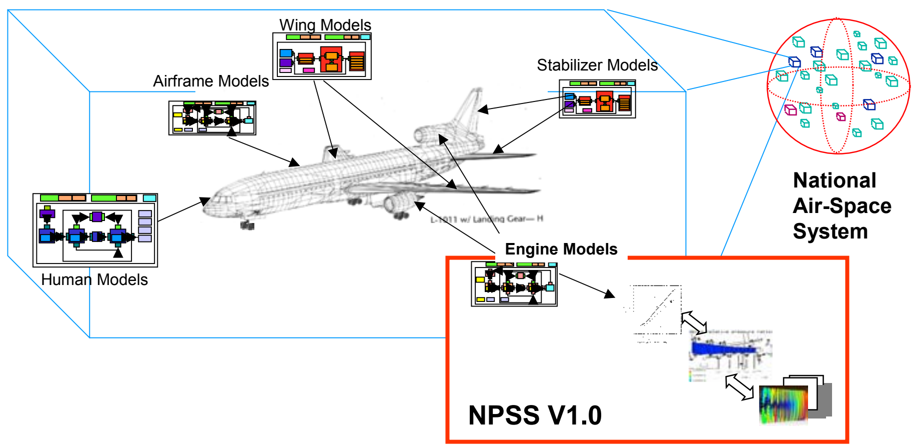
Across Multiple Grids (IPG and San Diego Supercomputer Center)



Aviation Safety

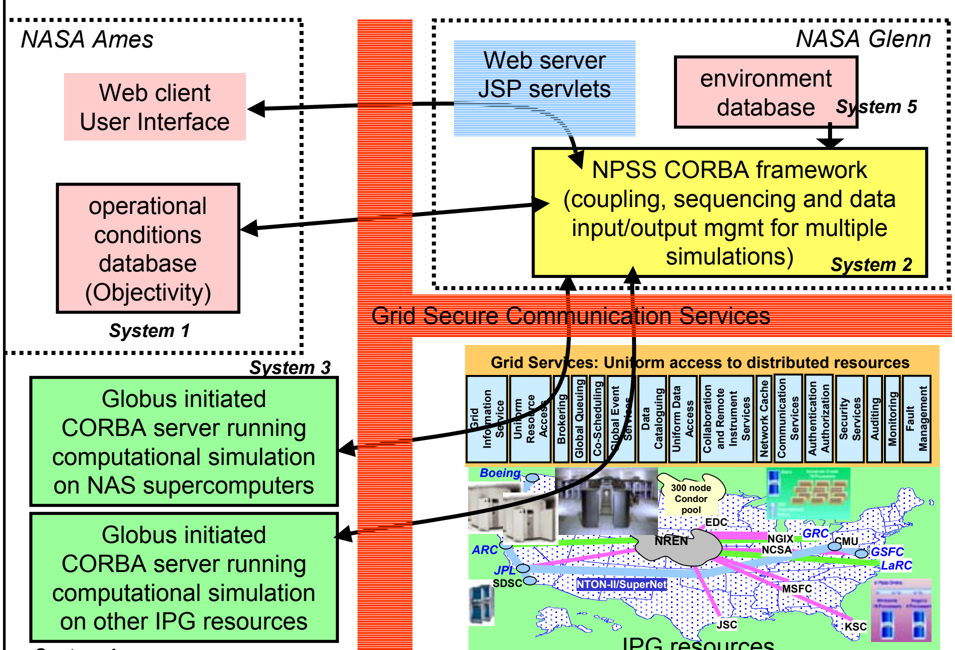
- Multiple sub-systems, e.g. a CFD model for a wing operating at NASA Ames and a turbo-machine model operating at NASA Glenn, are combined using GRC's NPSS (Numerical Propulsion System Simulation) application framework that manages the interactions of multiple models and uses IPG services to coordinate computing and data storage systems across NASA.

Multidisciplinary Problem Example
Aviation Safety



A wing CFD model and a turbo-machine model are combined as one of the first steps toward whole aircraft simulation in operational environments.

Multi-Disciplinary Simulation *Across Many Remote Systems*



These Examples Illustrate Grid Successes

- Standardized access to multi-institutional resources
- A common security approach and infrastructure
- Persistent services (Globus job management) that are used to *instantiate and run application frameworks on an as-needed basis*
  - CORBA
  - CONDOR job manager ("Glide-in")
  - Agent systems / servers (data mining example)

This allows users great flexibility in building their applications in the framework of their choice. They *do not have to rely on that framework being provided as a persistent service* on all of the computing systems where they need to run – they can instantiate their own environment using persistent Grid services.

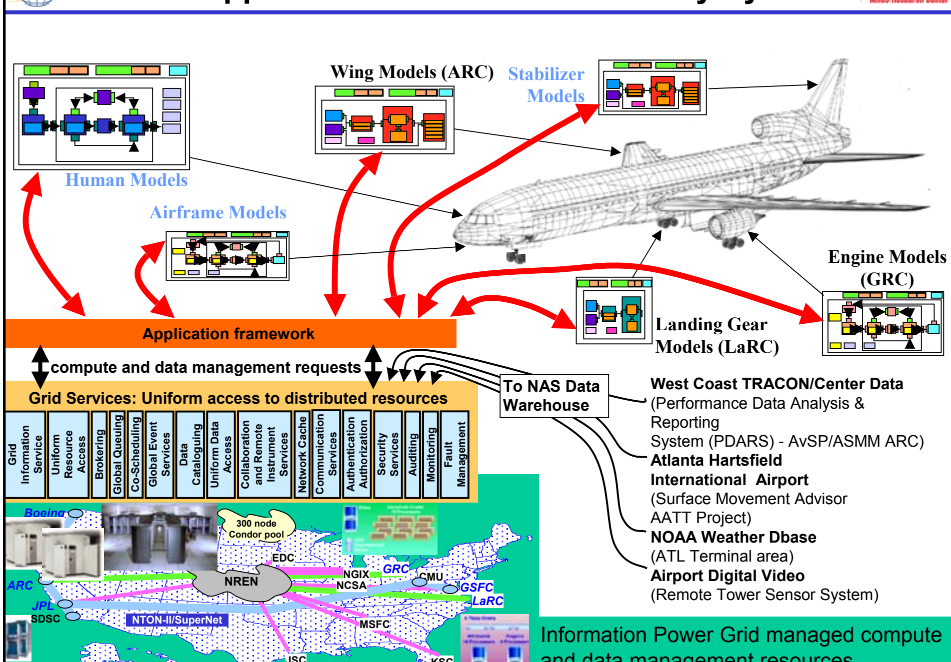
Information Power Grid Vision

- The IPG vision is to use the Grid approach to revolutionize how computing is used in NASA's science and engineering by providing the *middleware services for routinely building large-scale, dynamically constructed, and transient, problem solving environments from distributed, heterogeneous resources*

Information Power Grid Vision

- This revolution will come about enabling the *routine use of distributed resources*.
  From this we expect to see fundamental changes in how scientists and engineers access powerful computing systems, large-scale data archives, scientific instruments, and collaboration tools.
- IPG will facilitate these changes by providing services that are *integrated with the user's work environment* and that provide uniform and highly capable access to NASA's computers, data, and instruments, regardless of the locations or exact nature of these resources

Grids Support for Future Aviation Safety Systems



The Evolution of Grids

- Grids are currently focused on resource access and management
  - This is a necessary first step to provide a uniform underpinning, but is not sufficient if we want to realize the potential of Grids for facilitating science and engineering
  - Unless an application already has a framework that hides the use of these low level services (which was the case in all of the examples above), the Grid is difficult for most users

The Evolution of Grids

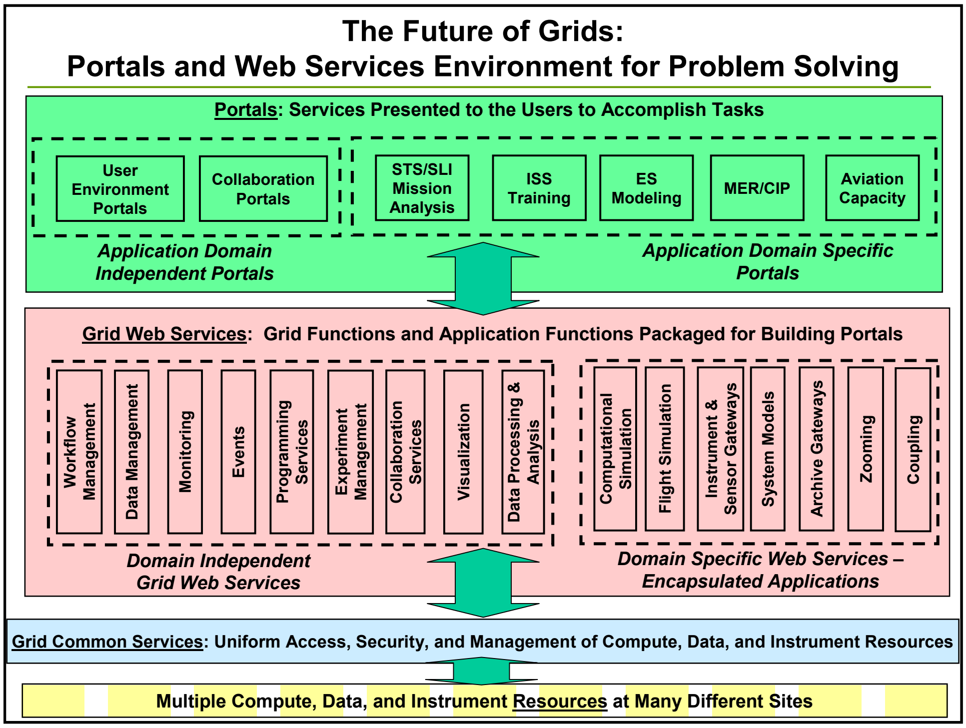
- Grids are evolving to a service oriented architecture
  - Users are primarily interested in "services" – something that performs a useful function, such as a particular type of simulation, or a broker that finds the "best" system to run a job
  - Even many Grid tool developers, such as those that develop application portals, are primarily interested in services – resource discovery, event management, user security credential management, etc.
- This evolution is going hand-in-hand with a large IT industry push to develop an integrated framework for Web services
- This is also what is necessary to address some of the current user complaints

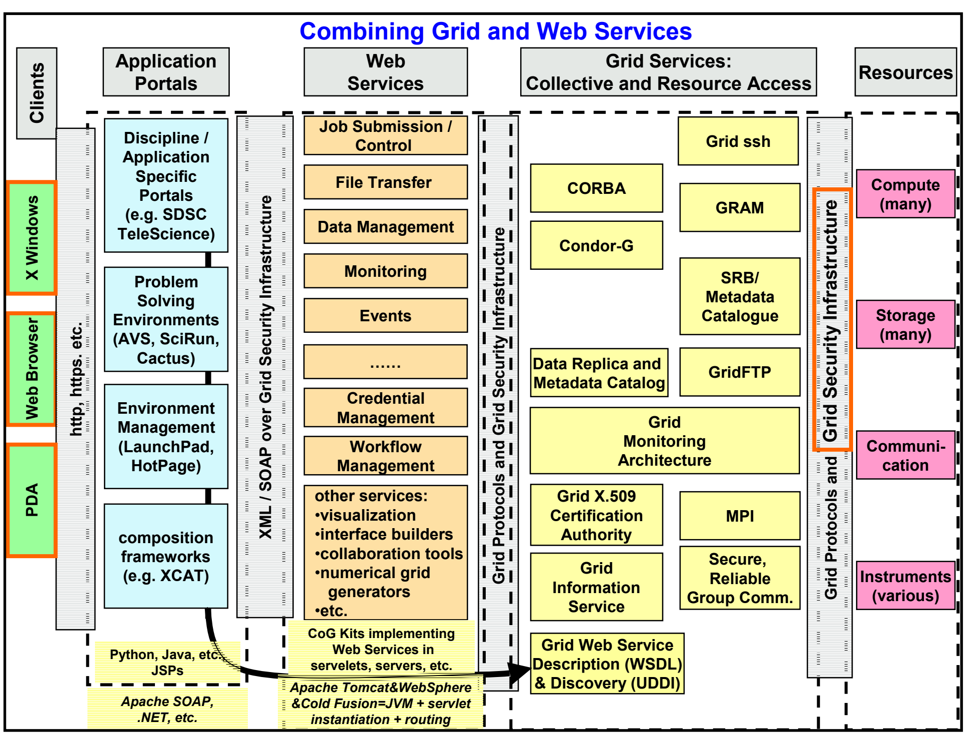
The Evolution of Grids

- Web services are a set of industry standards being developed and pushed by the major IT industry players (IBM, Microsoft, Sun, Compact, etc.)
  - A standard way to describe and discover Web accessible application components
  - A standard way to connect and interoperate these components
  - The IT industry expects that most, if not all, of it's applications to be packaged as Web services in the future

The Evolution of Grids

- Integrating Grids with Web services
  - Addresses several missing capabilities in the current Web Services approach (e.g. creating and managing job instances)
  - Makes the commercial investment in Web services tools – e.g. portal builders, graphical interface toolkits, etc. – available to the scientific community
  - Will provide for integrating commercial services with scientific and engineering applications and infrastructure
  - Currently a major thrust at the Global Grid Forum (See OGSI Working Group at www.gridforum.org)


The Future of Grids:
Portals and Web Services Environment for Problem Solving
Portals: Services Presented to the Users to Accomplish Tasks
User Environment Portals
Collaboration Portals
STS/SLI Mission Analysis
ISS Training
ES Modeling
MER/CIP
Aviation Capacity
Application Domain Independent Portals
Application Domain Specific Portals
Grid Web Services: Grid Functions and Application Functions Packaged for Building Portals
Workflow Management
Data Management
Monitoring
Events
Programming Services
Experiment Management
Collaboration Services
Visualization
Data Processing & Analysis
Computational Simulation
Flight Simulation
Instrument & Sensor Gateways
System Models
Archive Gateways
Zooming
Coupling
Domain Independent Grid Web Services
Domain Specific Web Services – Encapsulated Applications
Grid Common Services: Uniform Access, Security, and Management of Compute, Data, and Instrument Resources
Multiple Compute, Data, and Instrument Resources at Many Different Sites


Combining Grid and Web Services
Clients
Application Portals
Web Services
Grid Services: Collective and Resource Access
Resources
X Windows
Web Browser
PDA
http, https, etc.
Discipline / Application Specific Portals (e.g. SDSC TeleScience)
Problem Solving Environments (AVS, SciRun, Cactus)
Environment Management (LaunchPad, HotPage)
composition frameworks (e.g. XCAT)
Python, Java, etc. JSPs
Apache SOAP, .NET, etc.
XML / SOAP over Grid Security Infrastructure
Job Submission / Control
File Transfer
Data Management
Monitoring
Events
......
Credential Management
Workflow Management
other services:
•visualization
•interface builders
•collaboration tools
•numerical grid generators
•etc.
CoG Kits implementing Web Services in servelets, servers, etc.
Apache Tomcat&WebSphere &Cold Fusion=JVM + servlet instantiation + routing
Grid Protocols and Grid Security Infrastructure
CORBA
Condor-G
Grid ssh
GRAM
SRB/ Metadata Catalogue
Data Replica and Metadata Catalog
GridFTP
Grid Monitoring Architecture
Grid X.509 Certification Authority
MPI
Grid Information Service
Secure, Reliable Group Comm.
Grid Web Service Description (WSDL) & Discovery (UDDI)
Grid Protocols and Grid Security Infrastructure
Compute (many)
Storage (many)
Communi-cation
Instruments (various)

➢What is Missing?

- Knowledge Frameworks
  - From a problem description formulated by a scientist or engineer, be able to identify and automatically invoke appropriate operations on the computational components and datasets of the discipline area to “solve” the problem
- Science Portals/Problem-Solving Environments
  - General mechanisms and toolkits are needed for representing and manipulating the structure of the problem, and for *easily building portals* to instantiate this (e.g. with Web services)
- Workflow Management
  - Provide for description and subsequent control of the related steps and events that represent a “job.” A general approach is needed to provide a rule-based execution management system driven from published/subscribed global events (where the “events” represent process completion, file or other state creation, instrument turn-on, etc.)

What is Missing?

- Collaboration frameworks
  - Mechanisms for human control and sharing of all aspects of an executing workflow
- Global File System
  - Should provide Unix file semantics, be distributed, high performance, and use the Grid Security Infrastructure for authentication
- Application composing and dynamic execution
  - Need composition frameworks (e.g. IU XCAT) and dynamic object management in an environment of widely distributed resources (e.g. NSF GRADS)
- Monitoring / Global Events
  - Needed for all aspects of a running job (e.g. to support workflow mgmt and fault detection and recovery)

What is Missing?

- Authorization
  - Mechanisms to accommodate policy involving multiple stakeholders providing use-conditions on resources and user attributes in order to satisfy those use-conditions
- Dynamic construction of execution environments supporting complex distributed applications
  - Co-scheduling many resources to support transient science and engineering experiments that require combinations of instruments, compute systems, data archives, and network bandwidth at multiple locations (requires support by resource)
- Grid interfaces to existing commercial frameworks (e.g. MS DCOM and maybe IBM MQ)

Points to Take Away

1. Grids are a major international effort to define a common computing, data management, and collaboration infrastructure for science
2. NASA has a major role in Grids
3. NASA is heavily leveraging a lot of other work in Grids

- This talk is at grid.lbl.gov/~wej/Grids
- In a few days also at www.ipg.nasa.gov (About IPG -> Presentations)